

SkyIntelligent – 工业大数据科学云平台实践

南京天数信息科技有限公司 – 倪岭

- 动机 背景
- 大数 平台演进
- 设计理念
- 系统
- 底层 件 计算引 性能
- 中间层 容器平台 象
- 顶层 前 交 便利
- 工作流程
- 展望





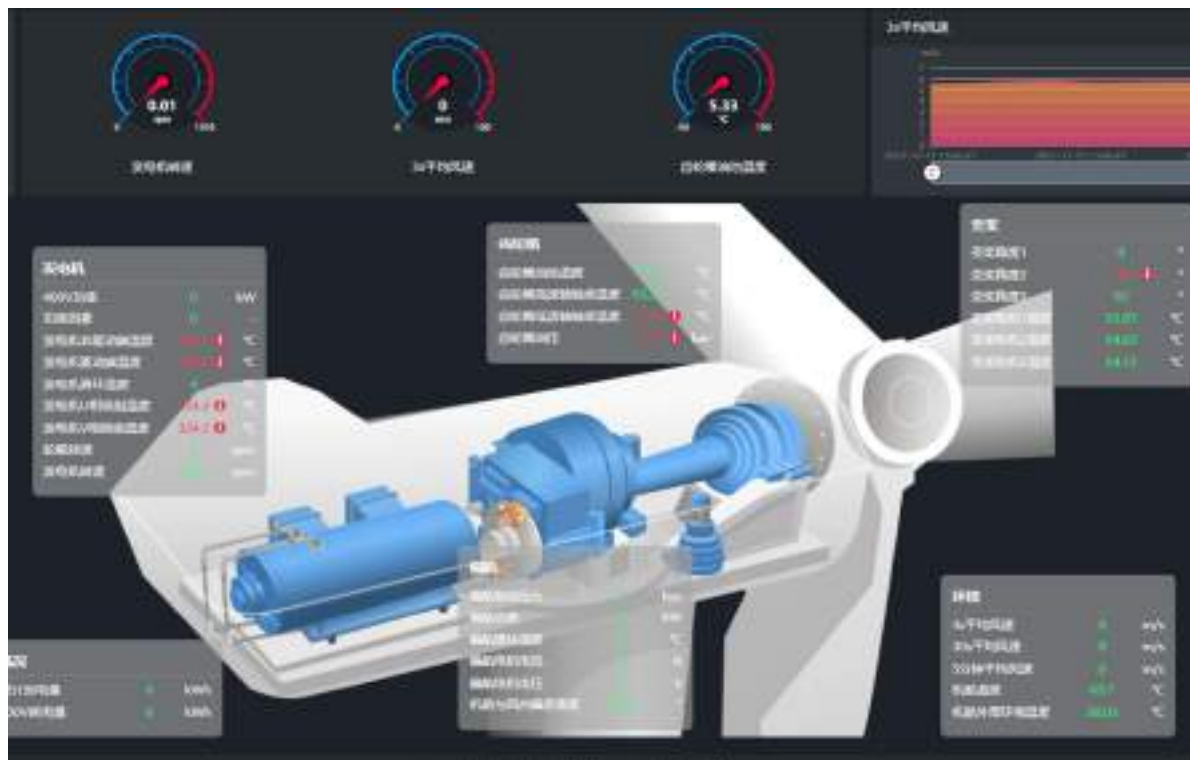
生产过程改进工艺改良



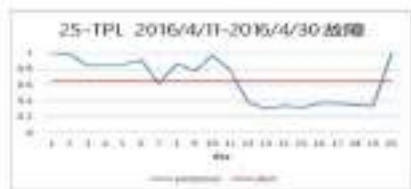
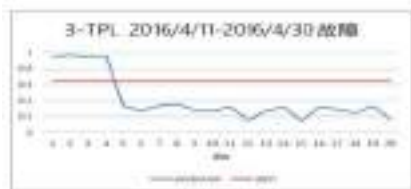
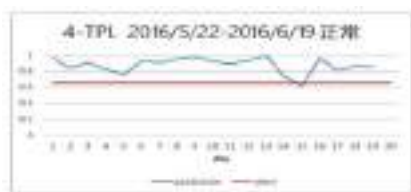
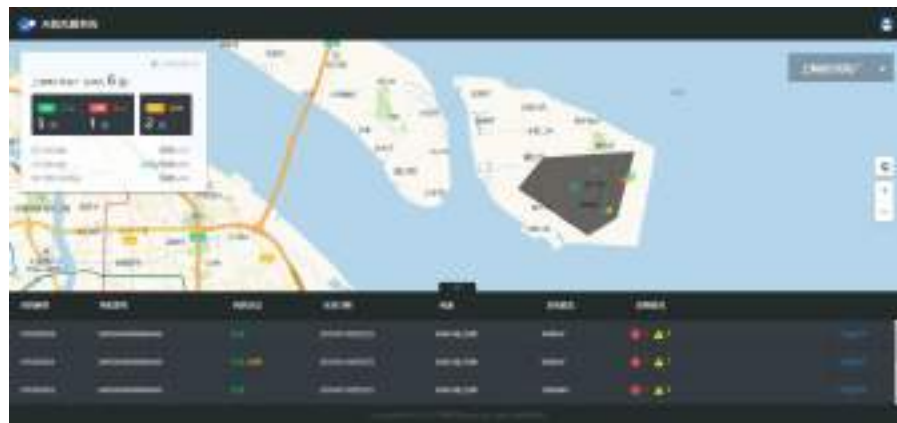
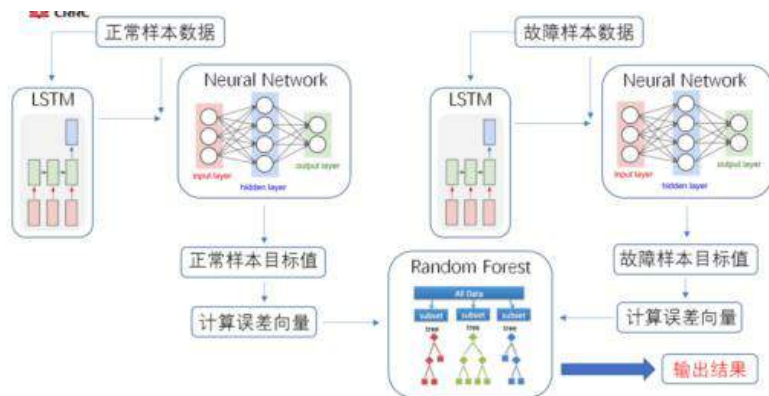
维护成本降低



更科学、高效的决策与管理



成本比较	之前	现在
齿轮箱故障	200万	10万
叶根螺栓断裂	120万	5万
叶片航拍视频分析 (1000台风机)	100万	5万
.....



Industry 4.0 framework and contributing digital technologies

• 数 分 求的演进

• 数 算法 计算力

• 企业私有云



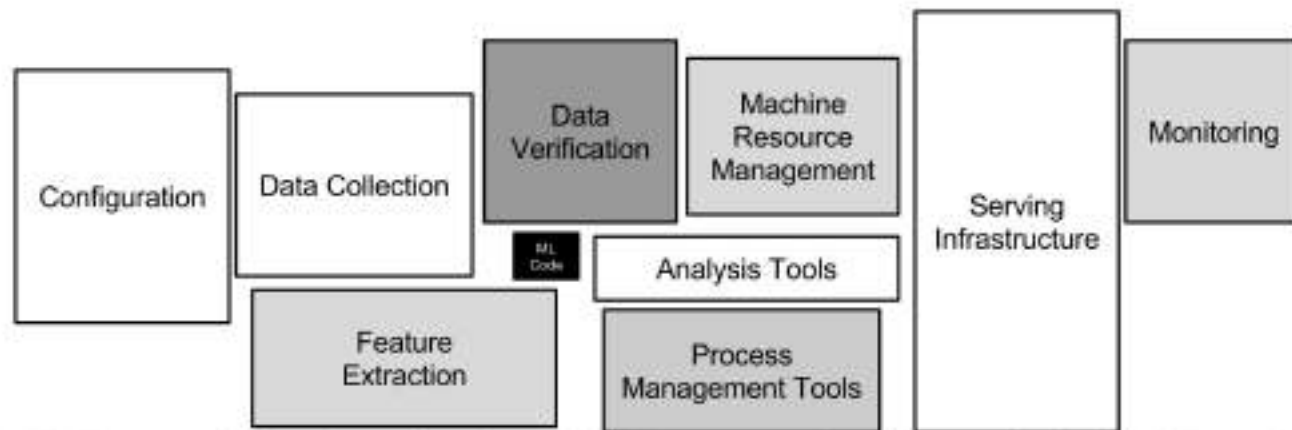


Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

- 机器学习 深度学习 不一求
- 大数据平台 机器学习平台 深度学习平台 服务平台
- 多平台
- 统一平台
- 企业级私有数 科学 平台

•

• 深度学习 化

•

— 高速 联

— 加速器

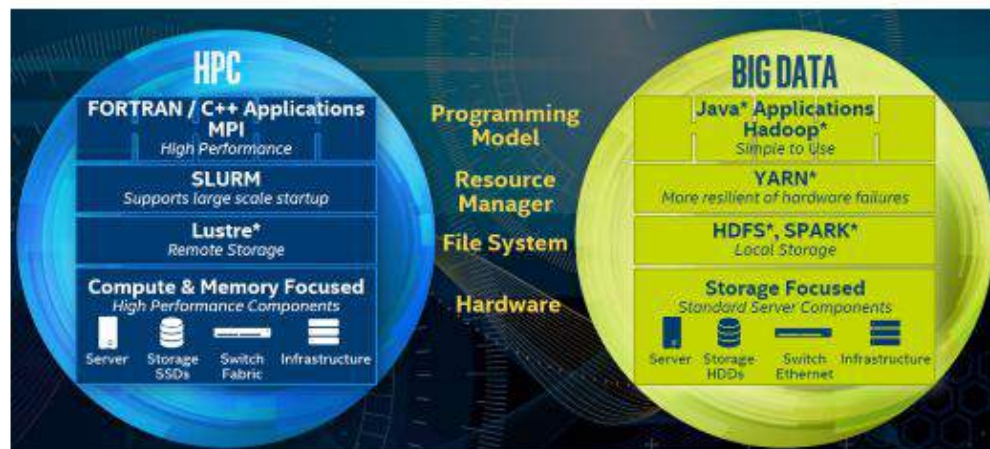
— 快速存储

—

— 合

n

THE CONVERGENCE OF HPC AND BIG DATA ANALYTIC SOLUTIONS



Other brands and names are the property of their respective owners.

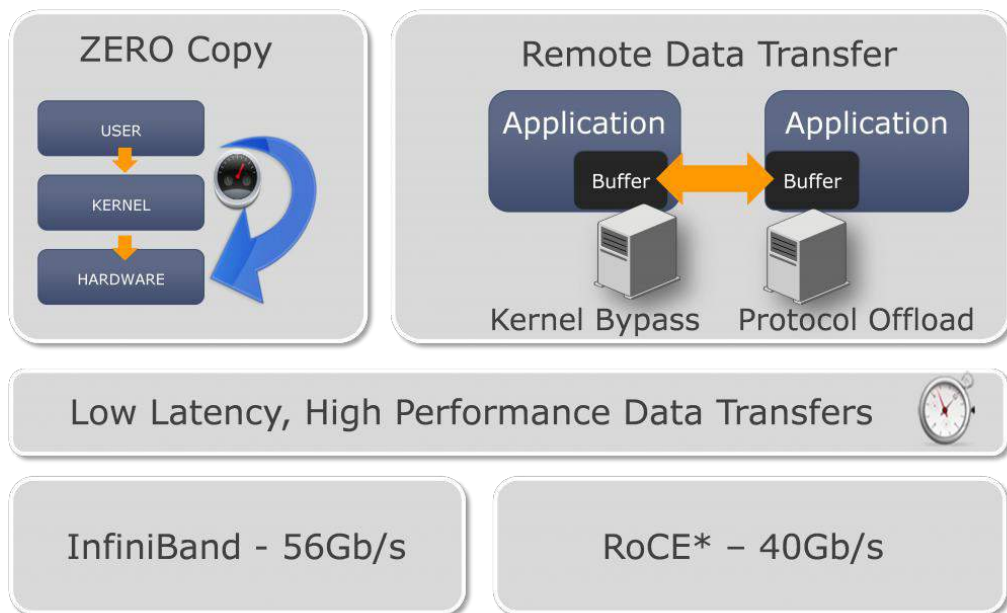
Sources: Daniel Reed and Jack Dongarra, Exascale Computing and Big Data in Communications of the ACM journal, July 2015 (Vol 58, No.7), and Intel analysis.

- 对接现有大数据平台软件资源重用
- 资源 拟化 弹性化
- 一式 全流程 模块化
- 高性能 高可用
- 多 户 合作 共
- 易扩展 可用户自定义



• 远程直接数 存取 的全面使用

—
—
—
—



* RDMA over Converged Ethernet

• 动态编译技术

—
—
—
—

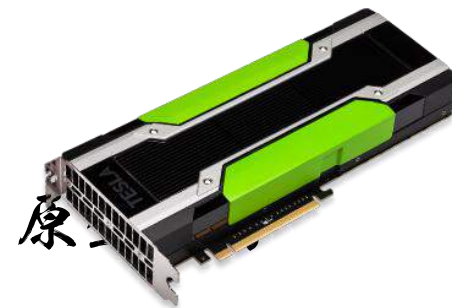
优化

• 混合存储优化

—

- 深度学习

—



- 数据库作为分布式中间件 前 交互式分

—

- 的 优化

—

- 的 优化

—

- 计算 通信重

- 计算步骤

内存 贝重

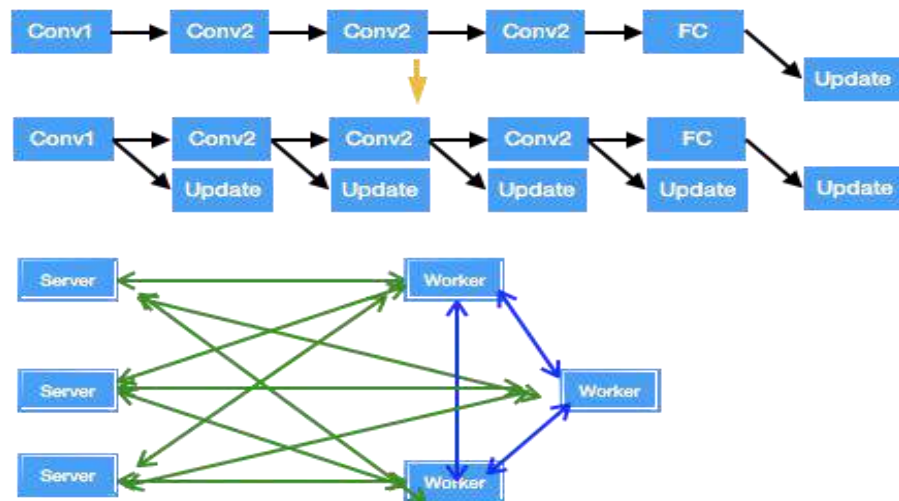
- 计算步骤

络数 交换 新重

- 混合通信方式

- 通过参数服务器传递小规模度

- 通过交换大规模度

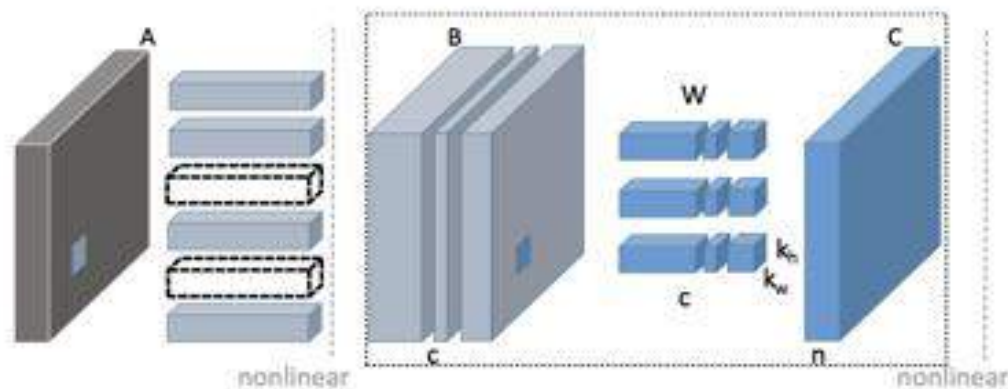


- 实现在 上的

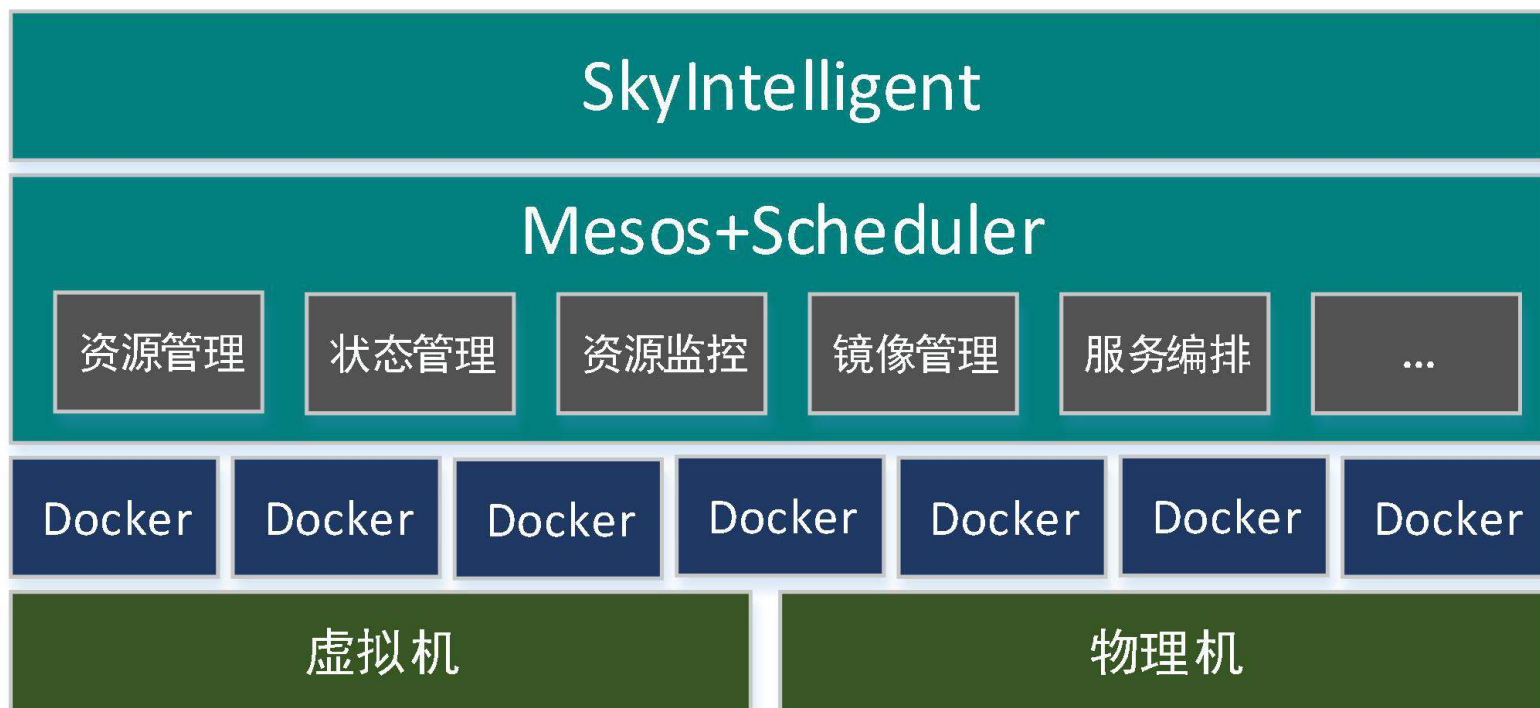
优化方法

• 通道剪枝

- 使用 L_1 范数 来选择 枝目标
- 使用最小二 来最小化重 误差
- 在不 率的前提下可以 倍的推
理速度提
- 同时 少 行时的内存 显存 用

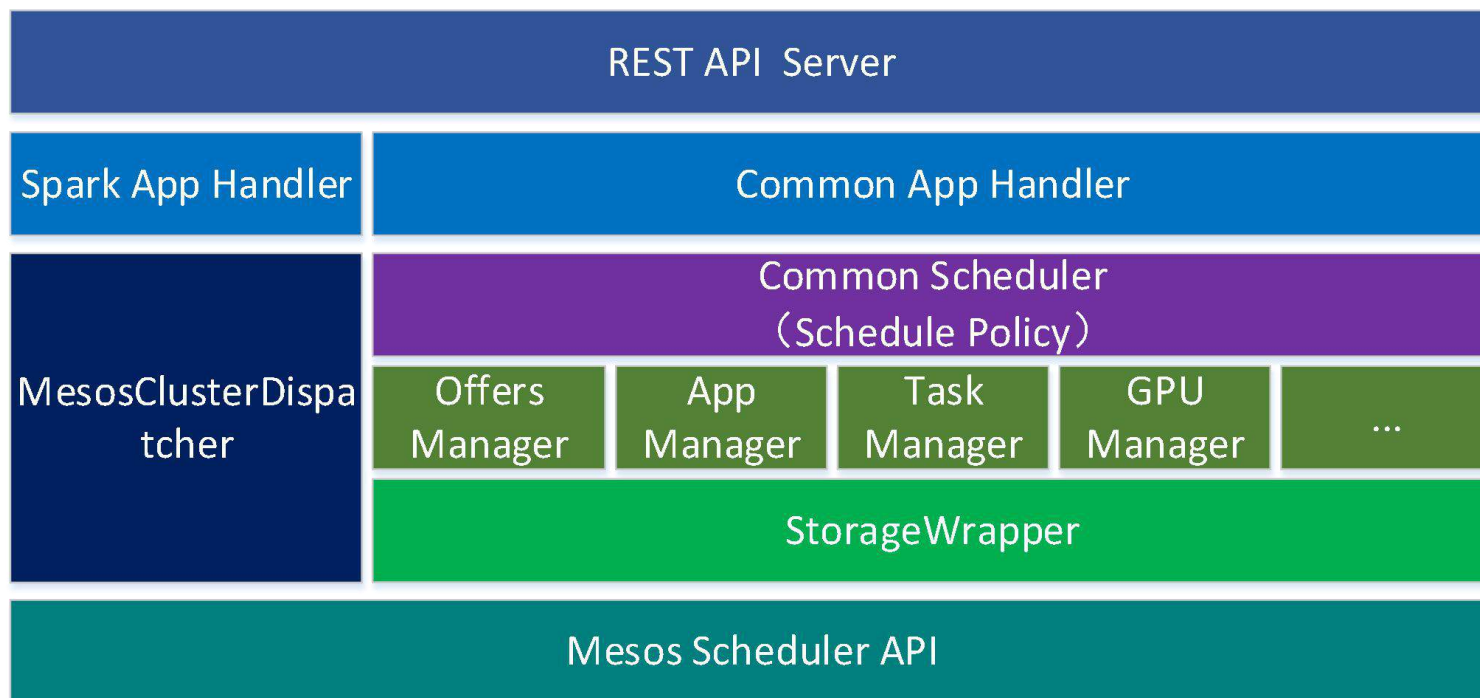


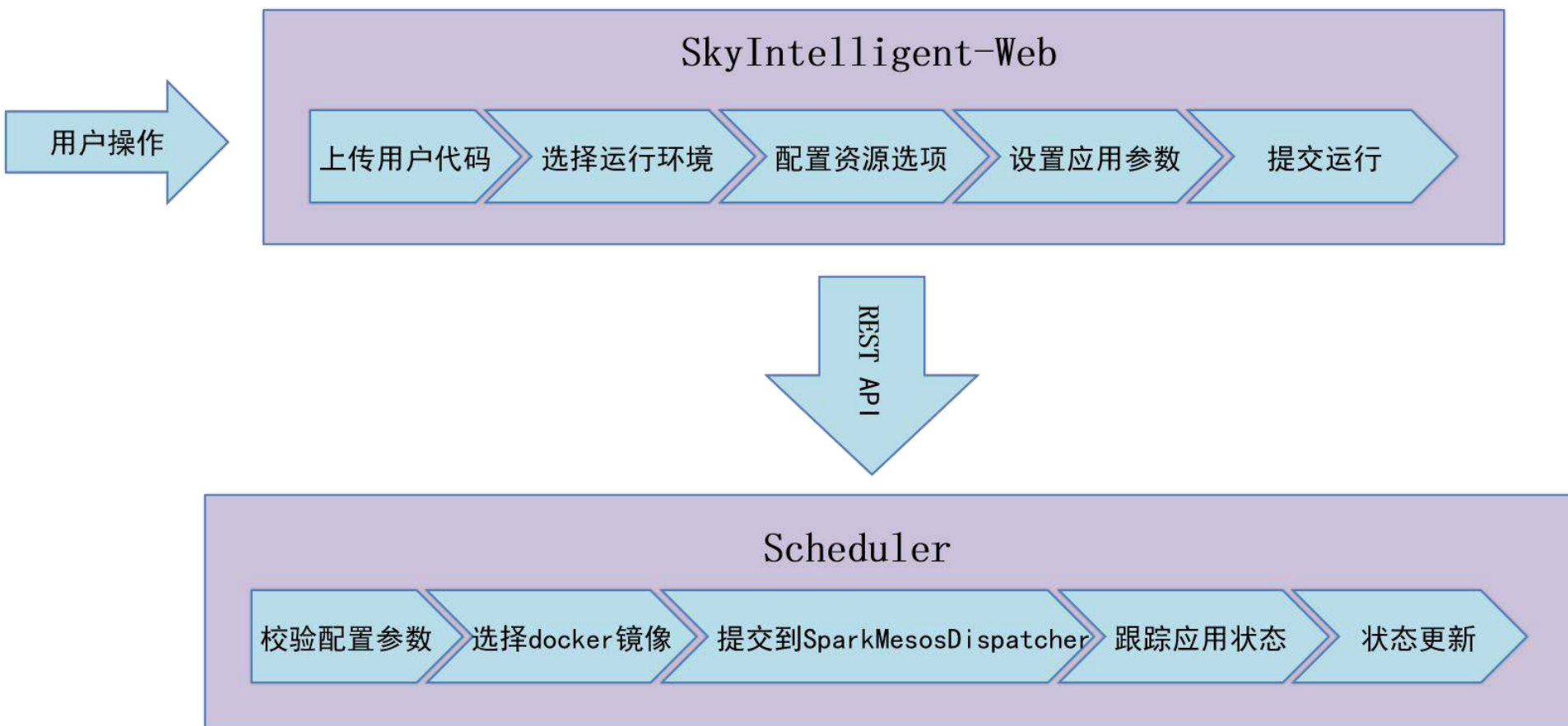
基 的容器化编 调度系统

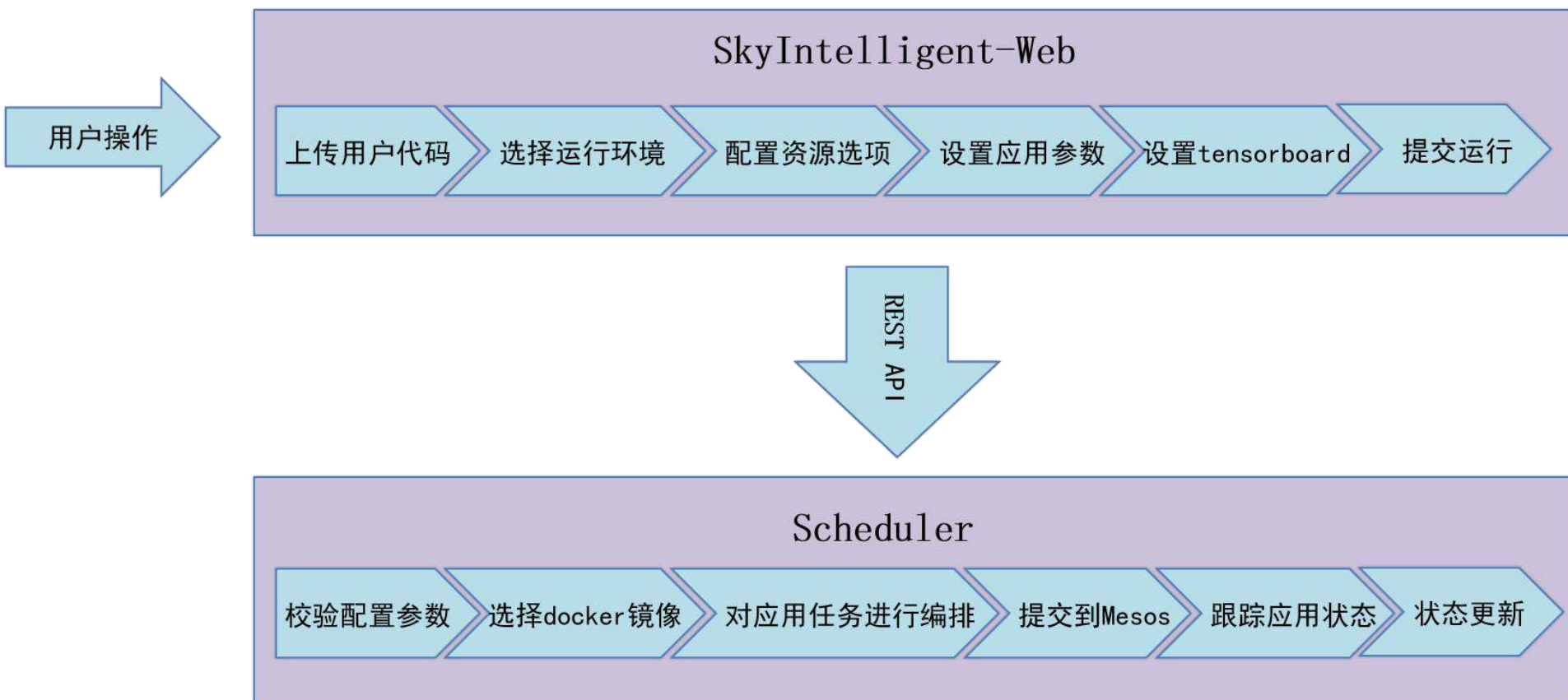


支持
深度学习

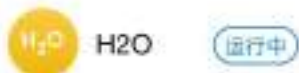
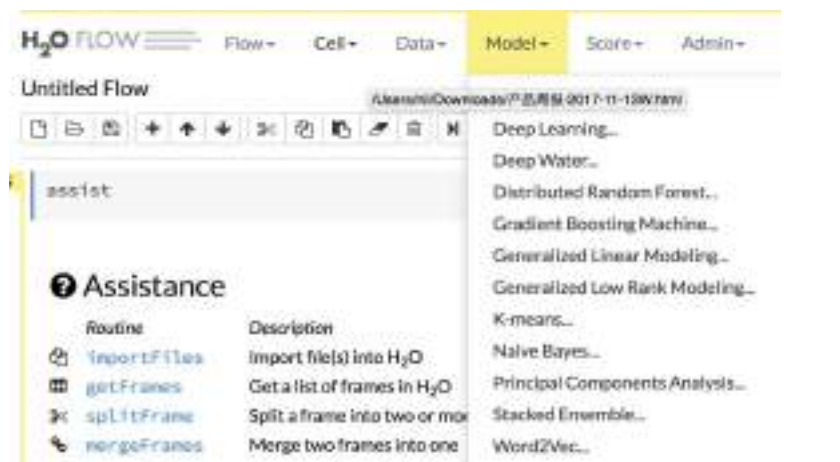
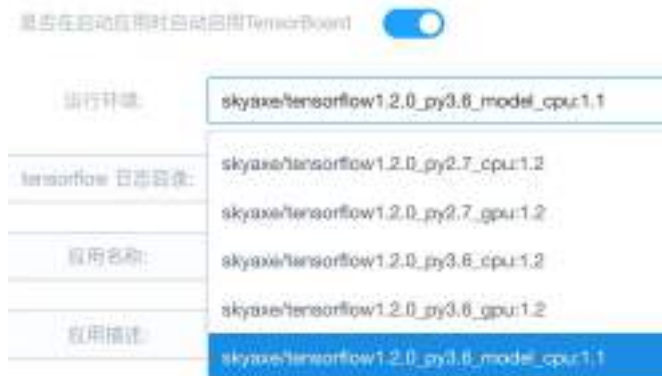
以 他较为流行的机器学习





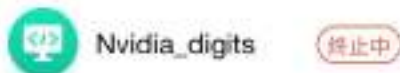


- 行环境 开发环境的用户自定义
- 支持在线定制 保存镜像文件
- 支持用户上传镜像 配置服务



打开工作空间 >

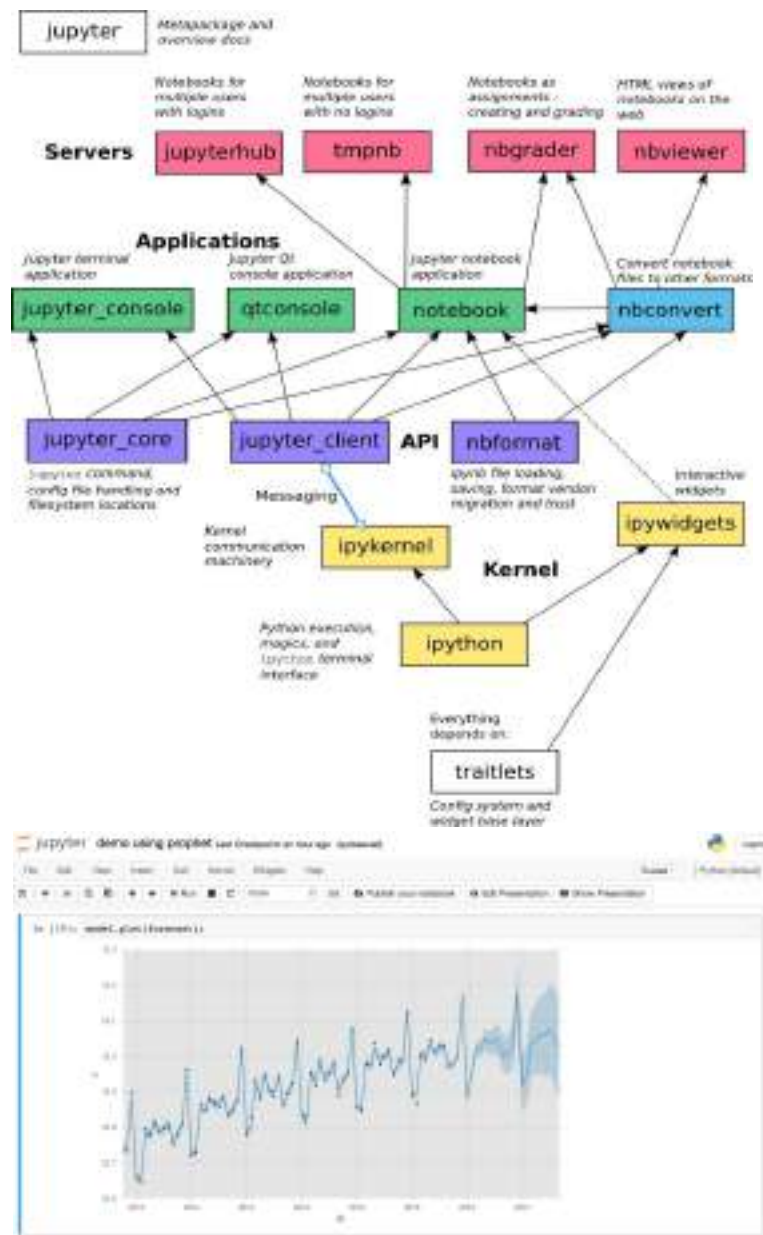
停止 详情

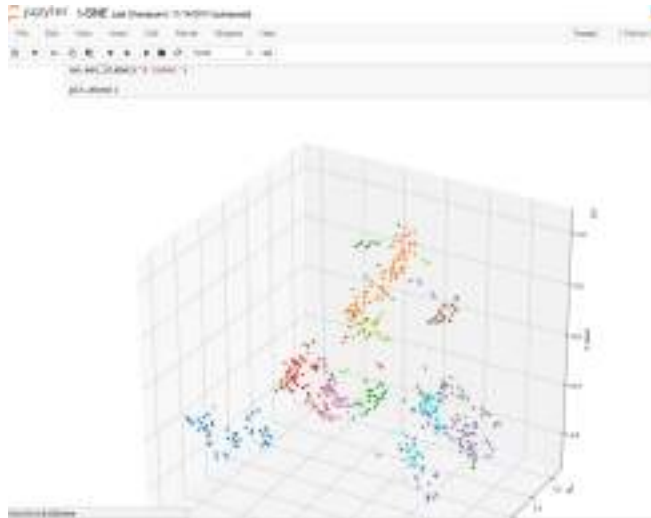


详情



- 项目 在开发开源软件 开放标准 用数十种编程语言的交互式计算服务
- 作为应用 提供创建 编辑 代码 方程式 图表 以 描述性文字 的文 可用 数 清洗 转换 统计建模 数 可视化 机器学习等方面
- 功能强大 大量的使用者 开发者 资源共
- 可视化交 式分 组件 富
- 可扩展





配置模型

选择模型:

LinearRegression

model_name:

model_name

☐ fit_intercept

☐ normalize

☐ copy_X

* y_fit

1

X:

1

0

1

确定 取消

Files Running Outputs Extensions

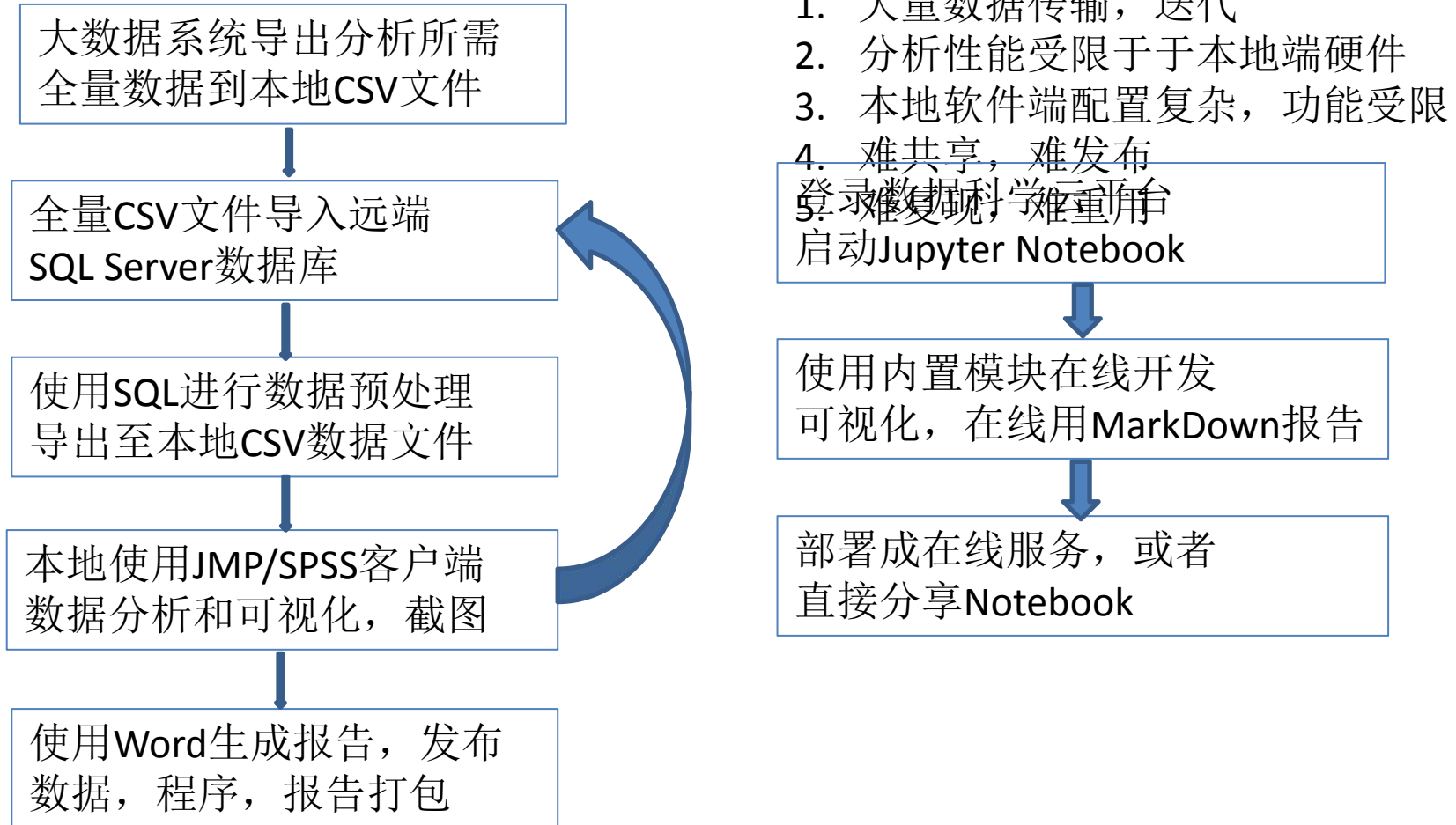
Configurable nbextensions

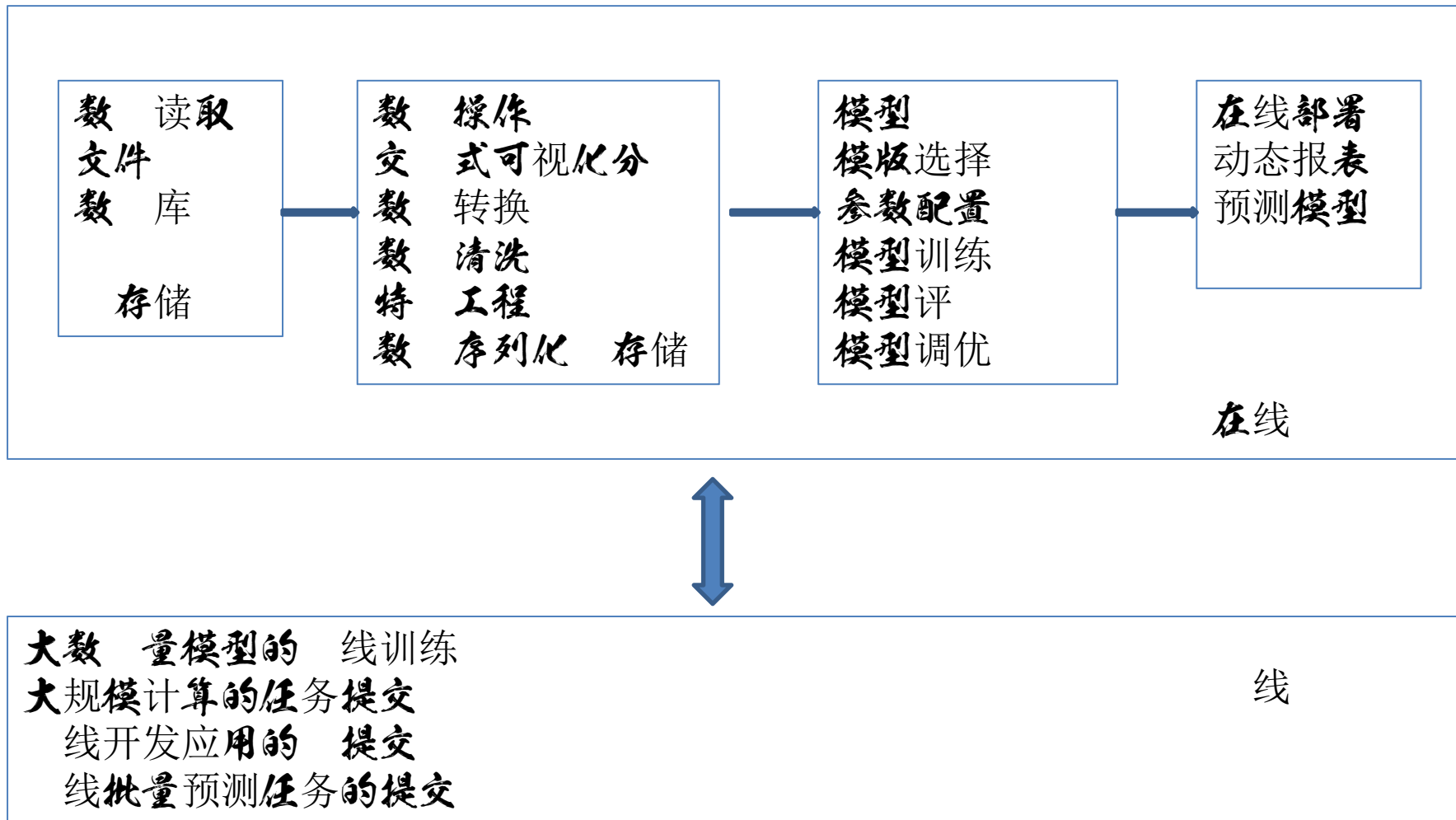
☒ disable configuration for nbextensions without explicit compatibility (they may break your notebook environment, but can be useful to share for nbextension development)

View by description, author, or tags

<input checked="" type="checkbox"/> JupyterLab environment for Jupyter	<input type="checkbox"/> JLab Converter	<input checked="" type="checkbox"/> Notebook	<input type="checkbox"/> Autoview
<input type="checkbox"/> AutoLoadTime	<input checked="" type="checkbox"/> Autocommit	<input checked="" type="checkbox"/> Notebook/nbextension	<input checked="" type="checkbox"/> Code-Test-Guide
<input type="checkbox"/> Code pretty	<input checked="" type="checkbox"/> Celltool	<input checked="" type="checkbox"/> CodeEditor in Editor	<input type="checkbox"/> CodeMirror mode extensions
<input type="checkbox"/> Collapsible Headings	<input type="checkbox"/> Command/Uncommand Filter	<input checked="" type="checkbox"/> console/nbextensions_help_menu	<input type="checkbox"/> ColorInspector
<input type="checkbox"/> Equation Auto Numbering	<input type="checkbox"/> ConsoleTime	<input type="checkbox"/> Exercise	<input type="checkbox"/> Exercise2
<input type="checkbox"/> Export Embedded HTML	<input type="checkbox"/> Fences	<input type="checkbox"/> Gedit	<input type="checkbox"/> Help-panel
<input type="checkbox"/> Hides Header	<input type="checkbox"/> Hide Input	<input type="checkbox"/> Highlight all	<input type="checkbox"/> Highlight selected word
<input type="checkbox"/> Highlighter	<input type="checkbox"/> Hideshow	<input type="checkbox"/> Initialization cells	<input checked="" type="checkbox"/> jupyter-jupyterlab/nbextension
<input type="checkbox"/> Keyboard shortcuts editor	<input checked="" type="checkbox"/> Launch JupyterLab	<input type="checkbox"/> Jupyter Output	<input type="checkbox"/> Move selected cells
<input type="checkbox"/> Navigation-Hotkeys	<input checked="" type="checkbox"/> Notebook/nbextension	<input checked="" type="checkbox"/> JupyterLab with notebook	<input type="checkbox"/> nbtimeline
<input type="checkbox"/> Modifi	<input type="checkbox"/> Preview	<input type="checkbox"/> Python Markdown	<input type="checkbox"/> RubricPanel
<input type="checkbox"/> Ruler	<input type="checkbox"/> Runbook	<input type="checkbox"/> Sketchpad	<input checked="" type="checkbox"/> ScrollDown
<input type="checkbox"/> Select CodeMirror Keymap	<input type="checkbox"/> SILENT Syntax	<input checked="" type="checkbox"/> Skip-Traceback	<input checked="" type="checkbox"/> SkyShots/nbextension
<input type="checkbox"/> Snippets	<input type="checkbox"/> Snippets Menu	<input type="checkbox"/> spellchecker	<input type="checkbox"/> Split Cells Notebook
<input checked="" type="checkbox"/> Table of Contents (2)	<input type="checkbox"/> Table_builder	<input type="checkbox"/> Toggle all line numbers	<input checked="" type="checkbox"/> Tree View
<input checked="" type="checkbox"/> Variable Inspector	<input checked="" type="checkbox"/> U.permite		

- 使用部署模型的应用协议
- 使用可视化库的应用快速部署可视化动态视图 报表
- 通过应用来转换部署服务
- 所有部署使用可选择的镜像 映射 部署均通过来协调 调度





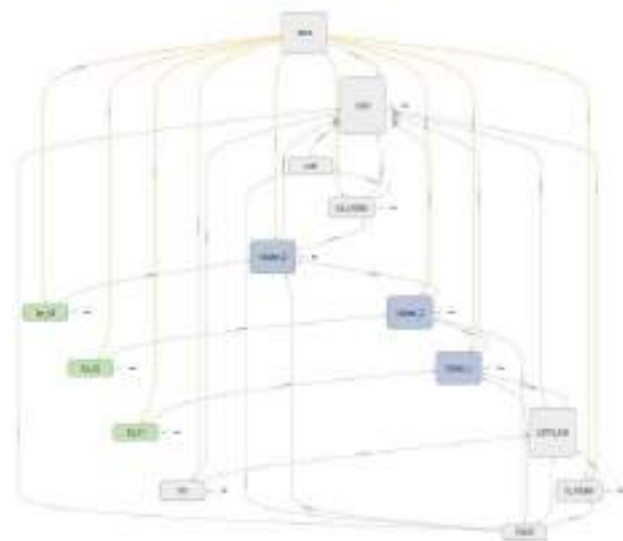
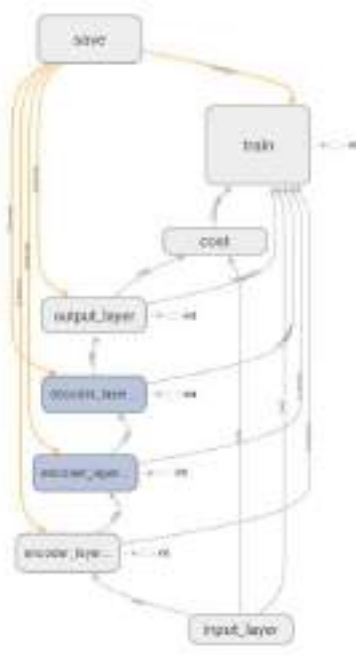
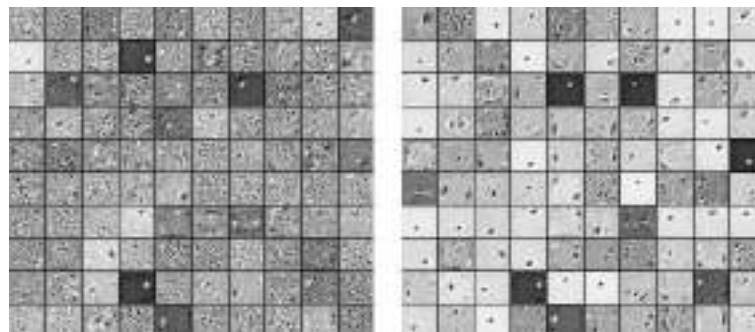
级别原始 数

数 格式转换 清洗 存储

时间序列分 探
常值
失值
合
统计标量
可视化探

预测模型探
动平均
时间序列分解
指数平
向量自回归
逻辑回归
支持向量机
神经网络

合为分钟级别数



降 自编码器进行滤波 络训练

- 在线 模块 模版的终 用户的自由定制
- 公有 部署 支持
- 多针对时序型 实时数 的优化
- 多内置的模型 行业模版
- 多内置交 式可视化组件
-
-



谢谢!

Ling.Ni@sky-data.cn

BDTC 2017 中国大数据技术大会
Big Data Technology Conference 2017